TAX RETURN NUMBERS TEND TO OBEY BENFORD'S LAW

By

Bruce Busta, C.P.A., Ph.D.
Department of Accounting
St. Cloud State University
St. Cloud, Minnesota 56301
(612) 255 - 3967

Richard Sundheim, Ph.D.
Associate Professor
Department of Mathematics and Statistics
St. Cloud State University
St. Cloud, Minnesota 56301
(612) 255 - 2239

St. Cloud State University
Center for Business Research
W93-106-94

22 April 1992

TAX RETURN NUMBERS TEND TO OBEY BENFORD'S LAW[1]

EXECUTIVE SUMMARY

Benford's Law predicts that the digits of naturally occurring numbers will follow a prescribed distribution. In previous research it has been found that street addresses, death rates, areas of rivers, population of cities, accounting measures of net income, dollar amounts on utilities bills, Fibonacci and Lucas numbers follow a Benford distribution.

This study examines tax return data from the 1982 and 1983 panel databases. It is discovered that these databases also tends to obey Benford's Law. The first digit of the tax return data closely follows a Benford distribution. While the second and third digits are reasonably close, there is an over-representation of zeros and fives in the second and third digit positions.

BACKGROUND

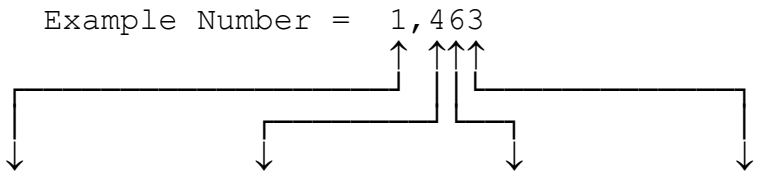
In 1938, Frank Benford published a paper which described a numerical phenomena which has come to be known as Benford's Law.[2]

---

[1] The research presented in this paper was conducted in the spring of 1992. This working paper was written in the summer of 1993. This working paper is generated from the authors' daily research journal. As a result, it presents the process followed by the authors. Thus, this paper differs from typical research papers in that hypothesis testing is not the primary goal of this paper. In addition to hypothesis testing, this paper reveals the various attempts, dead ends and findings encountered by the authors.

[2] See: Benford (1938). After 1938, it was uncovered that in

In that paper, Benford demonstrated that the digits of natural numbers are distributed in a predictable and specific pattern. For a large group of numbers, Benford's Law predicts that the first, second, third and fourth place digits (counting from the left) of the natural numbers will be distributed as follows:[3]



Digit	First Place	Second Place	Third Place	Fourth Place
0	-	.11968	.10178	.10018
1	.30103	.11389	.10138	.10014
2	.17609	.10882	.10097	.10010
3	.12494	.10433	.10057	.10006
4	.09691	.10031	.10018	.10002
5	.07918	.09668	.09979	.09998
6	.06695	.09337	.09940	.09994
7	.05799	.09035	.09902	.09990
8	.05115	.08757	.09864	.09986
9	.04576	.08500	.09827	.09982
Total	1.00	1.00	1.00	1.00

1881, Simon Newcomb described the same numerical phenomena that Benford described. As a consequence, Benford technically *rediscovered* the numerical phenomena that is named after him.

[3] The table is extracted from: McLaughlin, W. I., and S. A. Lundy. 1984. Digit Functions of Integer Sequences. Fibonacci Quarterly 22 (May): 109.

For example, if one opened his/her checkbook and determined how many checks began with 1, 2, 3, etc. One would expect to find that approximately 30 percent (.30103) of the checks would begin with 1, 18 percent (.17609) of the checks would begin with 2, 12 percent (.12494) would begin with 3, on so fourth. Similarly, 0 would be the value of the second digit for approximately 12 percent (.11968) of the checks, 1 would be the value of the second digit for approximately 11 percent (.11389) of the checks, and so on. For a complete explanation of the derivation of the above table or Benford's Law in general see: Sundheim and Busta (1993) or Nigrini (1991A).

In 1971, Wlodarski (1971) demonstrated that the first 100 Fibonacci and Lucas numbers closely followed a Benford distribution. Sentance (1973) extended and confirmed Wlodarski's work by observing the first 1,000 Fibonacci and Lucas numbers.

In this paper, Busta and Sundheim examine tax return data in order to determine if such data follows a Benford distribution.

DATA BASE

The tax return data used in this study was released by the Internal Revenue Service and made available by Ernst and Young through the Office of Tax Policy Research at the University of Michigan. The 1982 and 1983 panel databases are examined. These databases list Forms 1040 and 1040A for thousands of

taxpayers. For each individual 1040 or 1040A, there can be up to 190 data points. These individual data points represent virtually all the information found on schedules A, B, C, D, E and F of a tax return. See Appendix A for a detailed list of the variables on the 1982 and 1983 panel database.

To maintain taxpayer confidentiality, the IRS does not include either taxpayer names or social security numbers on the database. In addition, some data fields have been "blurred" or omitted.[4] The IRS also makes editing adjustments to insure the quality of the data. For example, director's fees may be reported on the "other income" line of the 1040. For consistency, the IRS would move this amount to the salaries and wages line of the Form 1040.

The numeric information on both data bases have been rounded to the four most significant digits (e.g., $14,371 = $14,370 and $228,867 = $228,900). Thus, only the first three digits of the numeric information are examined in this study.

METHOD AND RESULTS

The 1982 panel database is examined in great detail in this study. The 1983 panel database is briefly investigated. The method of analysis and results for the 1982 data are presented first; then the method of analysis and results are presented for

[4] The following fields have been modified to maintain confidentiality: Alimony Paid, Alimony Received, General Sales Tax, State and Local Taxes, Real Estate Taxes and Personal Property Taxes.

the 1983 tape.

A computer program was written which would analyze each variable field for each taxpayer in the database. When a non-zero number was encountered in one of the variable fields, the program would determine the value of the digit in the 1st, 2nd and 3rd digit position. This "digit count" was then aggregated to provide a total digit count for the entire database being examined. That is, the output from the computer program provided an aggregate digit count for each variable field selected for every taxpayer in the database.

For the initial analysis of the 1982 panel data base 144 fields were used to analyze how close tax return data follows Benford's Law. The 1982 panel database has 189 variable fields for each taxpayer. Fields E1_83 to E158_83 contain numeric data which would be expected to follow Benford's Law (This is actually 151 fields; fields E57, E91, E95, E105, E154, E155, and E156 are reserved). For this first analysis, fields E5, E36, E37, E38, E43, E45 and E153 were omitted because the field was a fixed amount (E5=dividend exclusion) or the field was an IRS calculated number (E45=marginal tax base) and was not specific to the taxpayer; such field are not expected to follow a Benford distribution. See Appendix A for a description of each of the fields and field numbers.

Table 1982-Attempt #1 shows the results of this first analysis. The results for the first and second digits show a digit pattern which closely follows a Benford distribution.

****

INSERT TABLE 1982-ATTEMPT #1 ABOUT HERE

****

In an effort to further refine the analysis of the 1982 tape, variable fields which were "blurred," calculated by the IRS (SOI), or a summation of several other fields were dropped. This resulted in the following fields being omitted: E1, E5, E6, E8, E15, E19, E21, E24, E25, E27, E33, E34, E36, E37, E38, E39, E40, E42, E43, E44, E45, E56, E59, E66, E74, E100, E120, E121, E123, E153. As a result 121[5] variable fields were re-analyzed for each of the 9235 taxpayers.

Because blurred and IRS calculated fields are not actual taxpayer generated numbers, they may not necessarily follow a Benford distribution. Consequently, they are omitted from the analysis. Fields which are the summation of several fields may not follow a Benford distribution, because when Benford "sets" (groups of numbers which follow a Benford distribution) are added the resulting set may not necessarily be a Benford set.

Table 1982-Attempt #2 shows the results of this investigation. This analysis reveals a closer conformity to a Benford set than Attempt #1. The first digit follows the Benford distribution very closely. The second and third digits

[5] Calculated as follows: E1 to E158 = 151 fields (7 reserved fields) less 30 judgementally dropped fields = 121.

are also close, except that there is a larger number of zeros and fives then the Benford distribution predicts.

****

INSERT TABLE 1982-ATTEMPT #2 ABOUT HERE

****

To further refine the analysis of the 1982 panel data base a correlation analysis was conducted for all variable fields. As expected, this revealed that many fields were highly correlated. For example, Salaries and Wages (E2) and EIC Salaries and Wages (E76) were perfectly correlated (Pearson Correlation = 1.00). Also, many of the medical deduction fields were highly correlated. This correlation is the result of identical numbers being on a taxpayer's tax return in several places. In order to eliminate this "double counting" in the digit analysis, such highly correlated fields were dropped. This resulted in the omission of variable fields E76, E83, E86, E88, E90, E98. Descriptive statistical analysis of the database reveal that many fields have less than 15 observations. Many variable fields have zero or only two data points. These small sample size variable fields were also dropped from the analysis. Thus, E30, E31, E53, E60, E72, E80, E141, E145, E148, E149 were omitted when conducting Attempt #3. As a result, 105 variable fields for 9235 taxpayers were used for the digit count in Attempt #3.

Table 1982-Attempt #3 surprisingly shows a distribution which is further away from a Benford distribution than Attempt #2. The difference between Table 1982-Attempt #2 and Table 1982-Attempt #3 is not large, but in the opposite direction from what was expected. This result leads the authors to conclude that the inclusion or exclusion of duplicate numbers in the analysis makes vary little difference in the overall distribution of the digits. This conclusion is confirmed later in this paper in Attempt #6 and #7. This conclusion implies that duplicate numbers which show up more than once on a tax return follow a Benford distribution.

****

INSERT TABLE 1982-ATTEMPT #3 ABOUT HERE

****

The above analysis has shown that the first digit closely follows a Benford set. While the second and third digits are reasonably close, there is an over-representation of zeros and fives. In an attempt to examine more closely the second and third digits an analysis was conducted for those numbers which were greater than or equal to 10,000. The rational for this analysis is that when numbers are estimated (e.g., charitable contributions) a two digit number it is likely to be 10, 20, 30, etc., or 15, 25, 35, etc. Thus causing an over-representation of zeros and fives in the second digit. A three digit number is

likely to be 100, 125, 150, 175, etc., again causing an over-representation. A four-digit number may also be subject to such estimation and concentration in the zeros and fives position. In essence, it assumed that such "lumpy" rounding will occur in the last two digits of a number; thus to examine three digits which are not subject to such lumpy rounding a five digit number must be examined, i.e., numbers greater than or equal to 10,000. See Christian and Gupta (1993) and Nigrini (1991B) for more discussion of rounding in the last two digit positions.

Table 1982-Attempt #4 shows the results of examining only those numbers on the database which are greater than or equal to 10,000. The variable fields used in this analysis are the same 105 fields used in Attempt #3.

The results of this strategy reveals a third digit distribution which very closely follows a Benford distribution. This is encouraging and adds support to the hypothesis that digits other than the first digit for tax return data follow a Benford distribution. The second digit does not follow a Benford distribution as closely as the other attempts. However, it does follow a distinctive pattern of over-representation in the low digits (0-4) and under-representation in the high digits (5-9). This systematic pattern may be a function of the truncation which occurs when only observing numbers greater than or equal to 10,000, rather than non-conformance with a Benford set. A new analysis with numbers which are only greater than or equal to 1,000 would help determine the cause of this systematic

pattern. Unfortunately, such an analysis was not conducted for this study. The first digit, as expected, does not follow a Benford set because of the truncation effect when only using numbers greater than or equal to 10,000.

****

INSERT TABLE 1982-ATTEMPT #4 ABOUT HERE

****

At this point, the authors felt that the over-representation of zeros and fives in the second and third digits may be the result of duplicate numbers on a taxpayer's tax return. For example, when looking at a taxpayer record it is not uncommon for Adjusted Gross Income (E1) to be the same number as Salary and Wages (E2), Tax Table Income Before Exemptions (E35), EIC Salary and Wages (E76) and EIC Earned Income (E77). Attempt #3 used a crude method (correlation analysis) to eliminate duplicate numbers. For Attempt #5, a sophisticated computer program was written to examine an individual taxpayer record and search for duplicate numbers. When a duplicate number was found, it was re-coded into a non-numeric form. Thus, when this computer program was finished an individual number on a taxpayer record only appeared once on that taxpayer record. This program to eliminate duplicate numbers required a large amount of computational time. The program was written in SPSS and ran on an older VAX mini-

computer. After nine hours of CPU time the program was aborted (The program was run at night and the actual elapsed time was greater than nine hours). At the completion of this run 9,181 taxpayer records had been processed. Because the computational time requirements were so large and the vast majority of the 9,235 taxpayer records had been voided of duplicate numbers the database of 9,181 records is used in the remainder of this study.

Attempt #5 uses a database of 9,181 taxpayer records. These individual tax returns contain no duplicate numbers. The following variable fields were dropped:

E31, E53, E80, E145, E149 - These are blank fields.

E153 - IRS taxpayer identification field.

E5, E36, E37 - Fields with fixed values.

Therefore, Attempt #5 analyzes 142[6] variable fields. Table 1982-Attempt #5 reports the results of this analysis. Again, the first digit closely follows the expected Benford distribution. Surprisingly, the second and third digits still are over-represented in the zero and fives digits. In fact, the elimination of the duplicate numbers had very little impact on the distribution of any of the digits.

****

INSERT TABLE 1982-ATTEMPT #5 ABOUT HERE

[6] E1 to E158 = 158 Fields less 7 Reserved fields less 9 field described above = 142.

****

In a final attempt to measure the conformity of tax return data to a Benford distribution the same variable fields used in Attempt #2 are again used in Attempt #6. The only difference between the two attempts is the database used. In Attempt #6, the database of 9,181 taxpayers with no duplicate numbers is used. In Attempt #2 the database of 9,235 taxpayers with duplicate numbers was used.

Therefore, Attempt #6 analyzes 121[7] variables for 9,181 taxpayers. Table 1982-Attempt #6 presents the results. These results show the first digit strongly conforming to a Benford distribution. The second digit follows the expected Benford distribution, except that the number of observed zeros and fives is greater than expected. The third closely follows a Benford set, except that the number of zeros is over-represented. This is consistent with the other attempts and again supports the contention that duplicate numbers do not effect the overall results.

Direct comparison of Attempt #2 and Attempt #6 is possible since the same exact variables are involved. Comparing Phi for the different attempts, one can see that the elimination of the duplicate numbers (Attempt #6) made the first digit slightly further away from a Benford distribution, while the second and

[7] Fields E1, E5, E6, E8, E15, E19, E21, E24, E25, E27, E33, E34, E36, E37, E38, E39, E40, E42, E43, E44, E45, E56, E59, E66, E74, E100, E120, E121, E123, E153 were dropped for Attempt #6.

third digits moved slightly closer to a Benford distribution.[8]

****

INSERT TABLE 1982-ATTEMPT #6 ABOUT HERE

****

In order to verify that the above results are not specific to the 1982 panel database, an analysis was made of the 1983 panel database.[9] For this investigation, 142 variable fields were used to analyze how close tax return data follows Benford's Law. The 1983 panel database has 190 fields for each taxpayer. Fields E1_83 to E152_83 contain numeric data which would be expected to follow Benford's Law (This is actually 148 fields; fields E20, E91, E95, and E105 are reserved). For this analysis fields E5, E36, E37, E38, E43 and E45 were omitted because the field was a fixed amount (E5=dividend exclusion) or the field was an IRS calculated number (E45=marginal tax base) and was not specific to the taxpayer. See Appendix A for a description of each of the fields and field numbers.

---

[8] It is not possible to test whether the small differences in Phi are statistically significant. This is one of the drawbacks of this statistic.

[9] Sub-sets of the 1982 panel data base were also analyzed. The first 100, 500 and 2,500 taxpayers records were investigated as well as a random sample of 100, 500 and 2,500 taxpayer records. The results for all of these analysis were virtually the same and similar to the results for the entire 1982 data base. Thus, supporting the assertion that the results are robust notwithstanding the sample size, variables selected or data base used. The results of these sub-set analysis can be obtained by contacting the authors.

The one key difference between the 1982 panel database and the 1983 is the increased number of taxpayers on the 1983 database. The 1983 tape contains 19,185 taxpayer records. Because of the size of this database the results presented below required 4.9 hours of CPU time on a VAX mini-computer.

The analysis of the 1983 database are presented on Table 1983-Attempt #1. These results are very similar to those found while using the 1982 database. The first digit conforms very nicely to Benford's Law, while the second and third digits are also close, the zero and five digit positions are over-represented. The overall degree of conformity is very close to that found with the 1982 database. As a result, we conclude that the results are not sensitive to the database used; rather the results are representative of tax return data in general.

****

INSERT TABLE 1983-ATTEMPT #1 ABOUT HERE

****

CONCLUSIONS

The results of the various attempts to analyze the 1982 and 1983 panel databases reveal that tax return data does tend to follow a Benford distribution. The first digit follows the expected distribution very closely. While the second and third digits are reasonably close, there is a greater occurrence of zeros and fives in the second and third position than predicted

by Benford's Law.

The inclusion or exclusion of different variable fields did not make a great deal of difference in the overall distribution of the digits. That is, the results for all the seven Attempts are about the same. This implies that the databases are quite robust in terms of their conformity with Benford's Law.

Interestingly, Attempts # 5 and #6 demonstrated that the inclusion or exclusion of duplicate numbers from the same taxpayer record makes very little difference in the overall distribution of the digits. This allows one to conclude, that duplicate numbers within these databases are numbers which follow a Benford distribution. When a duplicate number occurs within a taxpayer recorded, intuitively it would seem that a digit would become over-represented; however, this does not occur because in a different taxpayer record, a different duplicate number must "offset" the other duplicate number. With the overall result being that the entire database follows a Benford distribution.

Attempt #4 specifically examined the third digit on the 1982 database. This endeavor showed that if the affect of rounding can be decreased the digit tends to more closely follow a Benford distribution. Explicitly, the over-representation of zeros and fives is almost eliminated.

In summary, the aggregate results of this study, permits the conclusion that tax return data does tend to follow a Benford distribution.

ERRATA

28 January 1994

There is an error on page 6 of this paper. It is stated that Attempt #2 is calculated with 121 variable fields. This is incorrect; actually 115 variable fields were used. In Attempt #2 the following fields were omitted: E1, E5, E6, E8, E15, E19, E21, E24, E25, E27, E33, E34, E36, E37, E38, E39, E40, E42, E43, E44, E45, E56, E59, E66, E74, E100, E120, E121, E123, E153 and E76, E83, E86, E88, E90, E98 (See page 7). [E1 to E158 = 158 - 7 reserved - the 36 omitted fields listed above = 115]

Correspondingly, there is an error on page 12 and footnote 7 of this paper. It is stated that Attempt #6 is calculated with 121 variable fields. This is incorrect; actually 115 variable fields were used. In Attempt #6 the following fields were omitted: E1, E5, E6, E8, E15, E19, E21, E24, E25, E27, E33, E34, E36, E37, E38, E39, E40, E42, E43, E44, E45, E56, E59, E66, E74, E100, E120, E121, E123, E153 and E76, E83, E86, E88, E90, E98 (See page 7). [E1 to E158 = 158 - 7 reserved - the 36 omitted fields listed above = 115]

Consequently, all references to 121 variable fields in this paper and its tables should be changed to 115.

REFERENCES

Benford, F. 1938. The law of anomalous numbers. Proceedings of the American Philosophical Society Vol. 78 No. 4 (March 31): 551 - 572.

Christian, C., and S. Gupta. 1993. New evidence on "secondary evasion." The Journal of the American Taxation Association 15 (Spring): 72 - 93.

Nigrini, M. September 1991A. Benford's Law: A Literature Review and Taxation Application. Working Paper, Department of Accounting, University of Cincinnati, Cincinnati, Ohio.

Nigrini, M. 1991B. An assessment of the behavioral and revenue implications of individual tax tables. Working Paper, University of Cincinnati, Cincinnati, Ohio.

Sentance, W. A. 1973. A further analysis of benford's law. Fibonacci Quarterly 11: 490 - 494.

Sundheim, R. and B. Busta. 1993. Fibonacci numbers tend to obey benford's law: an extension of wlodarski and sentance. Center for Business Research Working Paper No. xx-xx-xx, St. Cloud State University, St. Cloud, Minnesota.

Wlodarski, J. 1971. Fibonacci and lucas numbers tend to obey benford's law. Fibonacci Quarterly Vol. 9 No. 1: 87 - 88.

KEY TO ABBREVIATIONS AND STATISTICS PRESENTED IN TABLES

Maximum = This is the maximum difference of the absolute value of the difference between the observed and expected values.

MAD = Mean Absolute Deviation = This is the absolute value of the difference between the observed and expected values summed and divided by the number of items summed. This statistic is a measure of the error between the observed and expected values. Each error is weighted equally.

MSE = Mean Squared Error = This is the difference between the observed and expected values squared, then summed and divided by the number of items summed. This statistic is a measure of the error between the observed and expected values. This statistic differs from the MAD in that large errors are weighted more heavily then smaller errors.

Phi = This is a statistic used to standardize the Chi-squared statistic. This is necessary when unequal sample sizes are used in the calculation of Chi-squared. Phi is calculated by taking the square root of the total Chi-squared divided by the sample size. In this study, a Phi close to zero indicates a smaller difference between the observed and expected values.

Contingency Coefficient = This is a second statistic used to standardize the Chi-squared statistic, which is necessary when unequal sample sizes are used in the calculation of Chi-squared. The Contingency Coefficient is calculated by taking the square root of the total Chi-squared divided by the addition of the sample size and the total Chi-squared. In this study, a Contingency Coefficient close to zero indicates a smaller difference between the observed and expected values.

TABLE 1982 - ATTEMPT #1

PERCENTAGE OF DIGITS IN THE 1ST POSITION FOR 144 VARIABLE FIELDS
ON THE 1982 PANEL DATA BASE FOR 9235 TAXPAYERS

First Digit	Observed Percentage	Expected Percentage	Difference in Percentage
1	29.839%	30.103%	-0.264%
2	19.562%	17.609%	1.953%
3	13.020%	12.494%	0.526%
4	9.599%	9.691%	-0.092%
5	7.673%	7.918%	-0.245%
6	6.240%	6.695%	-0.455%
7	5.241%	5.799%	-0.558%
8	4.692%	5.115%	-0.423%
9	4.132%	4.576%	-0.444%
Total	100.000%	100.000%	0.000%
Maximum			1.953%
MAD			0.551%
MSE			0.006%

TABLE 1982 - ATTEMPT #1

NUMBER OF DIGITS IN THE 1ST POSITION, 144 FIELDS, 1982 PANEL DATA BASE, 9235 TAXPAYERS

First Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
1	61759	62304	-545	-0.876%	4.776
2	40487	36446	4041	11.089%	448.163
3	26948	25859	1089	4.211%	45.865
4	19868	20058	-190	-0.945%	1.791
5	15880	16388	-508	-3.100%	15.745
6	12916	13857	-941	-6.789%	63.863
7	10848	12002	-1154	-9.617%	111.003
8	9712	10587	-875	-8.261%	72.249
9	8553	9471	-918	-9.693%	88.978
Total	206971	206971	0	-23.979%	
Maximum			4041	11.089%	
MAD			1140	6.064%	
MSE			2437317	0.503%	
Total Chi-Squared					852.433
Phi					0.06418
Contingency Coefficient					0.06404

TABLE 1982 - ATTEMPT #1

PERCENTAGE OF DIGITS IN THE 2ND POSITION FOR 144 VARIABLE FIELDS
ON THE 1982 PANEL DATA BASE FOR 9235 TAXPAYERS

Second Digit	Observed Percentage	Expected Percentage	Difference in Percentage
0	13.276%	11.968%	1.308%
1	10.741%	11.389%	-0.648%
2	10.819%	10.882%	-0.063%
3	9.967%	10.433%	-0.466%
4	9.529%	10.031%	-0.502%
5	11.293%	9.668%	1.625%
6	8.946%	9.337%	-0.391%
7	8.812%	9.035%	-0.223%
8	8.448%	8.757%	-0.309%
9	8.168%	8.500%	-0.332%
Total	100.000%	100.000%	0.000%
Maximum			1.625%
MAD			0.587%
MSE			0.006%

TABLE 1982 - ATTEMPT #1

NUMBER OF DIGITS IN THE 2ND POSITION, 144 FIELDS, 1982 PANEL DATA BASE, 9235 TAXPAYERS

Second Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
0	27300	24610	2690	10.931%	294.050
1	22087	23419	-1332	-5.689%	75.795
2	22248	22377	-129	-0.575%	0.741
3	20496	21453	-957	-4.463%	42.733
4	19594	20627	-1033	-5.007%	51.718
5	23222	19880	3342	16.808%	561.671
6	18396	19200	-804	-4.186%	33.648
7	18121	18579	-458	-2.464%	11.279
8	17372	18007	-635	-3.527%	22.400
9	16795	17479	-684	-3.911%	26.739
Total	205631	205631	0	-2.084%	
Maximum			3342	16.808%	
MAD			1206	5.756%	
MSE			2390429	0.531%	
Total Chi-Squared					1120.774
Phi					0.07383
Contingency Coefficient					0.07363

TABLE 1982 - ATTEMPT #1

PERCENTAGE OF DIGITS IN THE 3RD POSITION FOR 144 VARIABLE FIELDS
ON THE 1982 PANEL DATA BASE FOR 9235 TAXPAYERS

Third Digit	Observed Percentage	Expected Percentage	Difference in Percentage
0	14.085%	10.178%	3.907%
1	9.315%	10.138%	-0.823%
2	9.736%	10.097%	-0.361%
3	9.310%	10.057%	-0.747%
4	9.914%	10.018%	-0.104%
5	10.069%	9.979%	0.090%
6	9.590%	9.940%	-0.350%
7	9.207%	9.902%	-0.695%
8	9.662%	9.864%	-0.202%
9	9.112%	9.827%	-0.715%
Total	100.000%	100.000%	0.000%
Maximum			3.907%
MAD			0.799%
MSE			0.018%

TABLE 1982 - ATTEMPT #1

NUMBER OF DIGITS IN THE 3RD POSITION, 144 FIELDS, 1982 PANEL DATA BASE, 9235 TAXPAYERS

Third Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
0	26841	19395	7446	38.391%	2858.625
1	17751	19319	-1568	-8.115%	127.229
2	18552	19241	-689	-3.579%	24.647
3	17740	19164	-1424	-7.433%	105.872
4	18892	19090	-198	-1.038%	2.056
5	19187	19016	171	0.900%	1.542
6	18275	18941	-666	-3.519%	23.450
7	17544	18869	-1325	-7.022%	93.050
8	18412	18797	-385	-2.046%	7.871
9	17364	18726	-1362	-7.274%	99.081
Total	190558	190558	0	-0.734%	
Maximum			7446	38.391%	
MAD			1523	7.932%	
MSE			6467597	1.728%	
Total Chi-Squared					3343.422
Phi					0.13246
Contingency Coefficient					0.13131

TABLE 1982 - ATTEMPT #2

PERCENTAGE OF DIGITS IN THE 1ST POSITION FOR 121 VARIABLE FIELDS
ON THE 1982 PANEL DATA BASE FOR 9235 TAXPAYERS

First Digit	Observed Percentage	Expected Percentage	Difference in Percentage
1	29.817%	30.103%	-0.286%
2	18.181%	17.609%	0.572%
3	13.139%	12.494%	0.645%
4	9.704%	9.691%	0.013%
5	7.984%	7.918%	0.066%
6	6.547%	6.695%	-0.148%
7	5.444%	5.799%	-0.355%
8	4.883%	5.115%	-0.232%
9	4.301%	4.576%	-0.275%
Total	100.000%	100.000%	0.000%
Maximum			0.645%
MAD			0.288%
MSE			0.001%

TABLE 1982 - ATTEMPT #2

NUMBER OF DIGITS IN THE 1ST POSITION, 121 FIELDS, 1982 PANEL DATA BASE, 9235 TAXPAYERS

First Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
1	33572	33894	-322	-0.950%	3.057
2	20470	19827	643	3.246%	20.886
3	14794	14067	727	5.165%	37.533
4	10926	10911	15	0.134%	0.020
5	8989	8915	74	0.829%	0.612
6	7372	7538	-166	-2.203%	3.660
7	6129	6529	-400	-6.130%	24.538
8	5498	5759	-261	-4.534%	11.840
9	4843	5152	-309	-6.002%	18.563
Total	112593	112593	0	-10.446%	
Maximum			727	6.130%	
MAD			324	3.244%	
MSE			155888	0.153%	
Total Chi-Squared					120.708
Phi					0.03274
Contingency Coefficient					0.03273

TABLE 1982 - ATTEMPT #2

PERCENTAGE OF DIGITS IN THE 2ND POSITION FOR 121 VARIABLE FIELDS
ON THE 1982 PANEL DATA BASE FOR 9235 TAXPAYERS

Second Digit	Observed Percentage	Expected Percentage	Difference in Percentage
0	13.376%	11.968%	1.408%
1	10.799%	11.389%	-0.590%
2	10.829%	10.882%	-0.053%
3	10.092%	10.433%	-0.341%
4	9.716%	10.031%	-0.315%
5	10.386%	9.668%	0.718%
6	9.068%	9.337%	-0.269%
7	8.917%	9.035%	-0.118%
8	8.576%	8.757%	-0.181%
9	8.241%	8.500%	-0.259%
Total	100.000%	100.000%	0.000%
Maximum			1.408%
MAD			0.425%
MSE			0.003%

TABLE 1982 - ATTEMPT #2

NUMBER OF DIGITS IN THE 2ND POSITION, 121 FIELDS, 1982 PANEL DATA BASE, 9235 TAXPAYERS

Second Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
0	14945	13371	1574	11.769%	185.196
1	12065	12724	-659	-5.183%	34.179
2	12099	12158	-59	-0.485%	0.287
3	11275	11656	-381	-3.272%	12.478
4	10855	11207	-352	-3.143%	11.070
5	11604	10802	802	7.428%	59.596
6	10131	10432	-301	-2.884%	8.677
7	9963	10094	-131	-1.302%	1.712
8	9582	9784	-202	-2.063%	4.164
9	9207	9497	-290	-3.051%	8.838
Total	111726	111726	0	-2.186%	
Maximum			1574	11.769%	
MAD			475	4.058%	
MSE			406042	0.265%	
Total Chi-Squared					326.195
Phi					0.05403
Contingency Coefficient					0.05395

TABLE 1982 - ATTEMPT #2

PERCENTAGE OF DIGITS IN THE 3RD POSITION FOR 121 VARIABLE FIELDS
ON THE 1982 PANEL DATA BASE FOR 9235 TAXPAYERS

Third Digit	Observed Percentage	Expected Percentage	Difference in Percentage
0	13.699%	10.178%	3.521%
1	9.395%	10.138%	-0.743%
2	9.810%	10.097%	-0.287%
3	9.390%	10.057%	-0.667%
4	9.843%	10.018%	-0.175%
5	10.070%	9.979%	0.091%
6	9.638%	9.940%	-0.302%
7	9.330%	9.902%	-0.572%
8	9.658%	9.864%	-0.206%
9	9.168%	9.827%	-0.659%
Total	100.000%	100.000%	0.000%
Maximum			3.521%
MAD			0.722%
MSE			0.014%

TABLE 1982 - ATTEMPT #2

NUMBER OF DIGITS IN THE 3RD POSITION, 121 FIELDS, 1982 PANEL DATA BASE, 9235 TAXPAYERS

Third Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
0	14029	10423	3606	34.590%	1247.152
1	9622	10383	-761	-7.325%	55.709
2	10047	10341	-294	-2.839%	8.333
3	9616	10300	-684	-6.637%	45.368
4	10080	10260	-180	-1.751%	3.145
5	10313	10220	93	0.913%	0.852
6	9870	10180	-310	-3.043%	9.425
7	9555	10141	-586	-5.777%	33.844
8	9891	10102	-211	-2.088%	4.404
9	9389	10064	-675	-6.707%	45.276
Total	102412	102412	0	-0.663%	
Maximum			3606	34.590%	
MAD			740	7.167%	
MSE			1511180	1.398%	
Total Chi-Squared					1453.508
Phi					0.11913
Contingency Coefficient					0.11830

TABLE 1982 - ATTEMPT #3

PERCENTAGE OF DIGITS IN THE 1ST POSITION FOR 105 VARIABLE FIELDS
ON THE 1982 PANEL DATA BASE FOR 9235 TAXPAYERS

First Digit	Observed Percentage	Expected Percentage	Difference in Percentage
1	29.815%	30.103%	-0.288%
2	18.180%	17.609%	0.571%
3	13.142%	12.494%	0.648%
4	9.705%	9.691%	0.014%
5	7.985%	7.918%	0.067%
6	6.546%	6.695%	-0.149%
7	5.444%	5.799%	-0.355%
8	4.883%	5.115%	-0.232%
9	4.301%	4.576%	-0.275%
Total	100.000%	100.000%	0.000%
Maximum			0.648%
MAD			0.289%
MSE			0.001%

TABLE 1982 - ATTEMPT #3

NUMBER OF DIGITS IN THE 1ST POSITION, 105 FIELDS, 1982 PANEL DATA BASE, 9235 TAXPAYERS

First Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
1	33562	33886	-324	-0.955%	3.093
2	20464	19822	642	3.240%	20.810
3	14793	14064	729	5.183%	37.788
4	10924	10909	15	0.140%	0.021
5	8988	8913	75	0.842%	0.632
6	7369	7536	-167	-2.220%	3.714
7	6128	6528	-400	-6.123%	24.474
8	5497	5758	-261	-4.529%	11.809
9	4841	5151	-310	-6.019%	18.659
Total	112566	112566	0	-10.441%	
Maximum			729	6.123%	
MAD			325	3.250%	
MSE			156273	0.154%	
Total Chi-Squared					120.999
Phi					0.03279
Contingency Coefficient					0.03277

TABLE 1982 - ATTEMPT #3

PERCENTAGE OF DIGITS IN THE 2ND POSITION FOR 105 VARIABLE FIELDS
ON THE 1982 PANEL DATA BASE FOR 9235 TAXPAYERS

Second Digit	Observed Percentage	Expected Percentage	Difference in Percentage
0	13.378%	11.968%	1.410%
1	10.795%	11.389%	-0.594%
2	10.827%	10.882%	-0.055%
3	10.091%	10.433%	-0.342%
4	9.717%	10.031%	-0.314%
5	10.386%	9.668%	0.718%
6	9.069%	9.337%	-0.268%
7	8.920%	9.035%	-0.115%
8	8.575%	8.757%	-0.182%
9	8.243%	8.500%	-0.257%
Total	100.000%	100.000%	0.000%
Maximum			1.410%
MAD			0.426%
MSE			0.003%

TABLE 1982 - ATTEMPT #3

NUMBER OF DIGITS IN THE 2ND POSITION, 105 FIELDS, 1982 PANEL DATA BASE, 9235 TAXPAYERS

Second Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
0	14943	13368	1575	11.781%	185.530
1	12058	12721	-663	-5.215%	34.595
2	12094	12155	-61	-0.503%	0.307
3	11271	11654	-383	-3.283%	12.558
4	10854	11205	-351	-3.128%	10.966
5	11601	10799	802	7.426%	59.552
6	10130	10429	-299	-2.870%	8.591
7	9963	10092	-129	-1.278%	1.649
8	9578	9781	-203	-2.080%	4.233
9	9207	9494	-287	-3.027%	8.701
Total	111699	111699	0	-2.178%	
Maximum			1575	11.781%	
MAD			475	4.059%	
MSE			406661	0.265%	
Total Chi-Squared					326.683
Phi					0.05408
Contingency Coefficient					0.05400

TABLE 1982 - ATTEMPT #3

PERCENTAGE OF DIGITS IN THE 3RD POSITION FOR 105 VARIABLE FIELDS
ON THE 1982 PANEL DATA BASE FOR 9235 TAXPAYERS

Third Digit	Observed Percentage	Expected Percentage	Difference in Percentage
0	13.701%	10.178%	3.523%
1	9.393%	10.138%	-0.745%
2	9.809%	10.097%	-0.288%
3	9.391%	10.057%	-0.666%
4	9.843%	10.018%	-0.175%
5	10.069%	9.979%	0.090%
6	9.637%	9.940%	-0.303%
7	9.332%	9.902%	-0.570%
8	9.660%	9.864%	-0.204%
9	9.165%	9.827%	-0.662%
Total	100.000%	100.000%	0.000%
Maximum			3.523%
MAD			0.723%
MSE			0.014%

TABLE 1982 - ATTEMPT #3

NUMBER OF DIGITS IN THE 3RD POSITION, 105 FIELDS, 1982 PANEL DATA BASE, 9235 TAXPAYERS

Third Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
0	14028	10421	3607	34.615%	1248.608
1	9617	10380	-763	-7.350%	56.070
2	10043	10338	-295	-2.853%	8.413
3	9615	10297	-682	-6.623%	45.166
4	10078	10257	-179	-1.745%	3.125
5	10309	10217	92	0.899%	0.827
6	9867	10177	-310	-3.048%	9.453
7	9555	10138	-583	-5.753%	33.555
8	9890	10099	-209	-2.073%	4.340
9	9384	10061	-677	-6.733%	45.616
Total	102386	102386	0	-0.664%	
Maximum			3607	34.615%	
MAD			740	7.169%	
MSE			1512530	1.400%	
Total Chi-Squared					1455.173
Phi					0.11922
Contingency Coefficient					0.11838

TABLE 1982 - ATTEMPT #4

PERCENTAGE OF DIGITS IN THE 1ST POSITION FOR 105 VARIABLE FIELDS
ON THE 1982 PANEL DATA BASE FOR 9235 TAXPAYERS

First Digit	Observed Percentage	Expected Percentage	Difference in Percentage
1	51.161%	30.103%	21.058%
2	25.267%	17.609%	7.658%
3	12.550%	12.494%	0.056%
4	5.295%	9.691%	-4.396%
5	2.564%	7.918%	-5.354%
6	1.387%	6.695%	-5.308%
7	0.804%	5.799%	-4.995%
8	0.497%	5.115%	-4.618%
9	0.475%	4.576%	-4.101%
Total	100.000%	100.000%	0.000%
Maximum			21.058%
MAD			6.394%
MSE			0.713%

TABLE 1982 - ATTEMPT #4

NUMBER OF DIGITS IN THE 1ST POSITION, 105 FIELDS, 1982 PANEL DATA BASE, 9235 TAXPAYERS

First Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
1	9478	5577	3901	69.952%	2728.895
2	4681	3262	1419	43.490%	617.020
3	2325	2315	10	0.448%	0.046
4	981	1795	-814	-45.359%	369.383
5	475	1467	-992	-67.619%	670.701
6	257	1240	-983	-79.279%	779.567
7	149	1074	-925	-86.131%	796.988
8	92	948	-856	-90.291%	772.537
9	88	848	-760	-89.620%	680.885
Total	18526	18526	0	-344.409%	
Maximum			3901	90.291%	
MAD			1184	63.576%	
MSE			2445681	48.114%	
Total Chi-Squared					7416.022
Phi					0.63270
Contingency Coefficient					0.53467

TABLE 1982 - ATTEMPT #4

PERCENTAGE OF DIGITS IN THE 2ND POSITION FOR 105 VARIABLE FIELDS
ON THE 1982 PANEL DATA BASE FOR 9235 TAXPAYERS

Second Digit	Observed Percentage	Expected Percentage	Difference in Percentage
0	14.164%	11.968%	2.196%
1	12.420%	11.389%	1.031%
2	11.810%	10.882%	0.928%
3	11.060%	10.433%	0.627%
4	10.331%	10.031%	0.300%
5	9.198%	9.668%	-0.470%
6	8.923%	9.337%	-0.414%
7	7.816%	9.035%	-1.219%
8	7.460%	8.757%	-1.297%
9	6.817%	8.500%	-1.683%
Total	100.000%	100.000%	0.000%
Maximum			2.196%
MAD			1.017%
MSE			0.014%

TABLE 1982 - ATTEMPT #4

NUMBER OF DIGITS IN THE 2ND POSITION, 105 FIELDS, 1982 PANEL DATA BASE, 9235 TAXPAYERS

Second Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
0	2624	2217	407	18.348%	74.641
1	2301	2110	191	9.056%	17.304
2	2188	2016	172	8.532%	14.675
3	2049	1933	116	6.011%	6.984
4	1914	1858	56	2.995%	1.667
5	1704	1791	-87	-4.863%	4.235
6	1653	1730	-77	-4.438%	3.407
7	1448	1674	-226	-13.492%	30.467
8	1382	1622	-240	-14.813%	35.600
9	1263	1575	-312	-19.795%	61.702
Total	18526	18526	0	-12.459%	
Maximum			407	19.795%	
MAD			188	10.234%	
MSE			46758	1.373%	
Total Chi-Squared					250.681
Phi					0.11632
Contingency Coefficient					0.11555

TABLE 1982 - ATTEMPT #4

PERCENTAGE OF DIGITS IN THE 3RD POSITION FOR 105 VARIABLE FIELDS
ON THE 1982 PANEL DATA BASE FOR 9235 TAXPAYERS

Third Digit	Observed Percentage	Expected Percentage	Difference in Percentage
0	11.012%	10.178%	0.834%
1	10.202%	10.138%	0.064%
2	9.694%	10.097%	-0.403%
3	9.981%	10.057%	-0.076%
4	10.321%	10.018%	0.303%
5	9.943%	9.979%	-0.036%
6	9.560%	9.940%	-0.380%
7	9.981%	9.902%	0.079%
8	9.905%	9.864%	0.041%
9	9.403%	9.827%	-0.424%
Total	100.000%	100.000%	0.000%
Maximum			0.834%
MAD			0.264%
MSE			0.001%

TABLE 1982 - ATTEMPT #4

NUMBER OF DIGITS IN THE 3RD POSITION, 105 FIELDS, 1982 PANEL DATA BASE, 9235 TAXPAYERS

Third Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
0	2040	1886	154	8.190%	12.647
1	1890	1878	12	0.630%	0.075
2	1796	1871	-75	-3.986%	2.973
3	1849	1863	-14	-0.760%	0.108
4	1912	1856	56	3.021%	1.694
5	1842	1849	-7	-0.363%	0.024
6	1771	1841	-70	-3.828%	2.698
7	1849	1834	15	0.793%	0.115
8	1835	1827	8	0.416%	0.032
9	1742	1821	-79	-4.315%	3.389
Total	18526	18526	0	-0.202%	
Maximum			154	8.190%	
MAD			49	2.630%	
MSE			4434	0.127%	
Total Chi-Squared					23.754
Phi					0.03581
Contingency Coefficient					0.03578

TABLE 1982 - ATTEMPT #5

PERCENTAGE OF DIGITS IN THE 1ST POSITION FOR 142 VARIABLE FIELDS
ON THE 1982 PANEL DATA BASE FOR 9181 TAXPAYERS

First Digit	Observed Percentage	Expected Percentage	Difference in Percentage
1	29.685%	30.103%	-0.418%
2	20.056%	17.609%	2.447%
3	13.186%	12.494%	0.692%
4	9.771%	9.691%	0.080%
5	7.491%	7.918%	-0.427%
6	6.079%	6.695%	-0.616%
7	5.144%	5.799%	-0.655%
8	4.492%	5.115%	-0.623%
9	4.095%	4.576%	-0.481%
Total	100.000%	100.000%	0.000%
Maximum			2.447%
MAD			0.715%
MSE			0.009%

TABLE 1982 - ATTEMPT #5

NUMBER OF DIGITS IN THE 1ST POSITION, 142 FIELDS, 1982 PANEL DATA BASE, 9181 TAXPAYERS

First Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
1	44269	44893	-624	-1.390%	8.671
2	29910	26260	3650	13.897%	507.189
3	19664	18632	1032	5.536%	57.112
4	14572	14452	120	0.828%	0.992
5	11172	11808	-636	-5.388%	34.276
6	9066	9984	-918	-9.198%	84.464
7	7672	8648	-976	-11.287%	110.173
8	6699	7628	-929	-12.179%	113.153
9	6107	6824	-717	-10.510%	75.382
Total	149131	149131	0	-29.689%	
Maximum			3650	13.897%	
MAD			1067	7.802%	
MSE			2040571	0.807%	
Total Chi-Squared					991.411
Phi					0.08153
Contingency Coefficient					0.08127

TABLE 1982 - ATTEMPT #5

PERCENTAGE OF DIGITS IN THE 2ND POSITION FOR 142 VARIABLE FIELDS
ON THE 1982 PANEL DATA BASE FOR 9181 TAXPAYERS

Second Digit	Observed Percentage	Expected Percentage	Difference in Percentage
0	13.077%	11.968%	1.109%
1	10.615%	11.389%	-0.774%
2	10.786%	10.882%	-0.096%
3	9.944%	10.433%	-0.489%
4	9.468%	10.031%	-0.563%
5	11.446%	9.668%	1.778%
6	9.040%	9.337%	-0.297%
7	8.834%	9.035%	-0.201%
8	8.549%	8.757%	-0.208%
9	8.241%	8.500%	-0.259%
Total	100.000%	100.000%	0.000%
Maximum			1.778%
MAD			0.577%
MSE			0.006%

TABLE 1982 - ATTEMPT #5

NUMBER OF DIGITS IN THE 2ND POSITION, 142 FIELDS, 1982 PANEL DATA BASE, 9181 TAXPAYERS

Second Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
0	19373	17730	1643	9.268%	152.278
1	15725	16872	-1147	-6.799%	77.992
2	15979	16121	-142	-0.881%	1.251
3	14731	15456	-725	-4.690%	33.995
4	14027	14860	-833	-5.608%	46.730
5	16956	14323	2633	18.387%	484.201
6	13392	13832	-440	-3.182%	14.009
7	13087	13385	-298	-2.225%	6.626
8	12665	12973	-308	-2.374%	7.311
9	12209	12592	-383	-3.043%	11.664
Total	148144	148144	0	-1.148%	
Maximum			2633	18.387%	
MAD			855	5.646%	
MSE			1271496	0.554%	
Total Chi-Squared					836.058
Phi					0.07512
Contingency Coefficient					0.07491

TABLE 1982 - ATTEMPT #5

PERCENTAGE OF DIGITS IN THE 3RD POSITION FOR 142 VARIABLE FIELDS
ON THE 1982 PANEL DATA BASE FOR 9181 TAXPAYERS

Third Digit	Observed Percentage	Expected Percentage	Difference in Percentage
0	13.609%	10.178%	3.431%
1	9.481%	10.138%	-0.657%
2	9.660%	10.097%	-0.437%
3	9.476%	10.057%	-0.581%
4	9.846%	10.018%	-0.172%
5	10.090%	9.979%	0.111%
6	9.518%	9.940%	-0.422%
7	9.371%	9.902%	-0.531%
8	9.656%	9.864%	-0.208%
9	9.293%	9.827%	-0.534%
Total	100.000%	100.000%	0.000%
Maximum			3.431%
MAD			0.708%
MSE			0.014%

TABLE 1982 - ATTEMPT #5

NUMBER OF DIGITS IN THE 3RD POSITION, 142 FIELDS, 1982 PANEL DATA BASE, 9181 TAXPAYERS

Third Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
0	18678	13969	4709	33.709%	1587.341
1	13012	13914	-902	-6.484%	58.499
2	13258	13858	-600	-4.329%	25.972
3	13006	13803	-797	-5.774%	46.023
4	13514	13750	-236	-1.713%	4.034
5	13848	13696	152	1.110%	1.687
6	13063	13642	-579	-4.247%	24.612
7	12862	13590	-728	-5.359%	39.029
8	13252	13538	-286	-2.114%	6.048
9	12755	13487	-732	-5.430%	39.767
Total	137248	137248	0	-0.631%	
Maximum			4709	33.709%	
MAD			972	7.027%	
MSE			2554585	1.315%	
Total Chi-Squared					1833.012
Phi					0.11557
Contingency Coefficient					0.11480

TABLE 1982 - ATTEMPT #6

PERCENTAGE OF DIGITS IN THE 1ST POSITION FOR 121 VARIABLE FIELDS
ON THE 1982 PANEL DATA BASE FOR 9181 TAXPAYERS

First Digit	Observed Percentage	Expected Percentage	Difference in Percentage
1	29.732%	30.103%	-0.371%
2	18.433%	17.609%	0.824%
3	13.436%	12.494%	0.942%
4	9.794%	9.691%	0.103%
5	7.903%	7.918%	-0.015%
6	6.417%	6.695%	-0.278%
7	5.382%	5.799%	-0.417%
8	4.659%	5.115%	-0.456%
9	4.244%	4.576%	-0.332%
Total	100.000%	100.000%	0.000%
Maximum			0.942%
MAD			0.415%
MSE			0.003%

TABLE 1982 - ATTEMPT #6

NUMBER OF DIGITS IN THE 1ST POSITION, 121 FIELDS, 1982 PANEL DATA BASE, 9181 TAXPAYERS

First Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
1	24055	24355	-300	-1.231%	3.691
2	14913	14247	666	4.678%	31.175
3	10870	10108	762	7.536%	57.402
4	7924	7841	83	1.065%	0.889
5	6394	6406	-12	-0.188%	0.023
6	5192	5417	-225	-4.146%	9.312
7	4354	4692	-338	-7.197%	24.304
8	3769	4138	-369	-8.924%	32.955
9	3434	3702	-268	-7.245%	19.431
Total	80905	80905	0	-15.653%	
Maximum			762	8.924%	
MAD			336	4.690%	
MSE			166019	0.314%	
Total Chi-Squared					179.182
Phi					0.04706
Contingency Coefficient					0.04701

TABLE 1982 - ATTEMPT #6

PERCENTAGE OF DIGITS IN THE 2ND POSITION FOR 121 VARIABLE FIELDS
ON THE 1982 PANEL DATA BASE FOR 9181 TAXPAYERS

Second Digit	Observed Percentage	Expected Percentage	Difference in Percentage
0	13.235%	11.968%	1.267%
1	10.654%	11.389%	-0.735%
2	10.872%	10.882%	-0.010%
3	10.081%	10.433%	-0.352%
4	9.722%	10.031%	-0.309%
5	10.431%	9.668%	0.763%
6	9.078%	9.337%	-0.259%
7	8.992%	9.035%	-0.043%
8	8.631%	8.757%	-0.126%
9	8.304%	8.500%	-0.196%
Total	100.000%	100.000%	0.000%
Maximum			1.267%
MAD			0.406%
MSE			0.003%

TABLE 1982 - ATTEMPT #6

NUMBER OF DIGITS IN THE 2ND POSITION, 121 FIELDS, 1982 PANEL DATA BASE, 9181 TAXPAYERS

Second Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
0	10618	9601	1017	10.590%	107.681
1	8547	9137	-590	-6.454%	38.062
2	8722	8730	-8	-0.091%	0.007
3	8087	8370	-283	-3.378%	9.553
4	7799	8047	-248	-3.085%	7.659
5	8368	7756	612	7.890%	48.282
6	7283	7491	-208	-2.770%	5.749
7	7214	7248	-34	-0.472%	0.162
8	6924	7025	-101	-1.441%	1.458
9	6662	6819	-157	-2.303%	3.617
Total	80224	80224	0	-1.516%	
Maximum			1017	10.590%	
MAD			326	3.848%	
MSE			197690	0.252%	
Total Chi-Squared					222.230
Phi					0.05263
Contingency Coefficient					0.05256

TABLE 1982 - ATTEMPT #6

PERCENTAGE OF DIGITS IN THE 3RD POSITION FOR 121 VARIABLE FIELDS
ON THE 1982 PANEL DATA BASE FOR 9181 TAXPAYERS

Third Digit	Observed Percentage	Expected Percentage	Difference in Percentage
0	13.307%	10.178%	3.129%
1	9.620%	10.138%	-0.518%
2	9.871%	10.097%	-0.226%
3	9.509%	10.057%	-0.548%
4	9.792%	10.018%	-0.226%
5	9.973%	9.979%	-0.006%
6	9.581%	9.940%	-0.359%
7	9.448%	9.902%	-0.454%
8	9.663%	9.864%	-0.201%
9	9.237%	9.827%	-0.590%
Total	100.000%	100.000%	0.000%
Maximum			3.129%
MAD			0.626%
MSE			0.011%

TABLE 1982 - ATTEMPT #6

NUMBER OF DIGITS IN THE 3RD POSITION, 121 FIELDS, 1982 PANEL DATA BASE, 9181 TAXPAYERS

Third Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
0	9748	7456	2292	30.739%	704.500
1	7047	7427	-380	-5.114%	19.422
2	7231	7397	-166	-2.241%	3.715
3	6966	7367	-401	-5.449%	21.876
4	7173	7339	-166	-2.260%	3.750
5	7306	7310	-4	-0.059%	0.003
6	7019	7282	-263	-3.608%	9.481
7	6921	7254	-333	-4.589%	15.278
8	7079	7226	-147	-2.035%	2.993
9	6767	7199	-432	-6.000%	25.920
Total	73257	73257	0	-0.618%	
Maximum			2292	30.739%	
MAD			458	6.210%	
MSE			600133	1.085%	
Total Chi-Squared					806.936
Phi					0.10495
Contingency Coefficient					0.10438

TABLE 1983 - ATTEMPT #1

PERCENTAGE OF DIGITS IN THE 1ST POSITION FOR 142 VARIABLE FIELDS
ON THE 1983 PANEL DATA BASE FOR 19185 TAXPAYERS

First Digit	Observed Percentage	Expected Percentage	Difference in Percentage
1	29.164%	30.103%	-0.939%
2	19.999%	17.609%	2.390%
3	13.170%	12.494%	0.676%
4	9.686%	9.691%	-0.005%
5	7.730%	7.918%	-0.188%
6	6.187%	6.695%	-0.508%
7	5.267%	5.799%	-0.532%
8	4.652%	5.115%	-0.463%
9	4.145%	4.576%	-0.431%
Total	100.000%	100.000%	0.000%
Maximum			2.390%
MAD			0.681%
MSE			0.009%

TABLE 1983 - ATTEMPT #1

NUMBER OF DIGITS IN THE 1ST POSITION, 142 FIELDS, 1983 PANEL DATA BASE, 19185 TAXPAYERS

First Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
1	127680	131792	-4112	-3.120%	128.287
2	87557	77093	10464	13.574%	1420.380
3	57659	54699	2960	5.411%	160.167
4	42404	42427	-23	-0.055%	0.013
5	33842	34665	-823	-2.375%	19.551
6	27086	29311	-2225	-7.591%	168.887
7	23061	25388	-2327	-9.166%	213.321
8	20367	22394	-2027	-9.050%	183.409
9	18147	20034	-1887	-9.418%	177.712
Total	437803	437803	0	-21.791%	
Maximum			10464	13.574%	
MAD			2983	6.640%	
MSE			17097880	0.601%	
Total Chi-Squared					2471.727
Phi					0.07514
Contingency Coefficient					0.07493

TABLE 1983 - ATTEMPT #1

PERCENTAGE OF DIGITS IN THE 2ND POSITION FOR 142 VARIABLE FIELDS
ON THE 1983 PANEL DATA BASE FOR 19185 TAXPAYERS

Second Digit	Observed Percentage	Expected Percentage	Difference in Percentage
0	13.364%	11.968%	1.396%
1	10.713%	11.389%	-0.676%
2	10.891%	10.882%	0.009%
3	9.954%	10.433%	-0.479%
4	9.909%	10.031%	-0.122%
5	10.594%	9.668%	0.926%
6	9.066%	9.337%	-0.271%
7	8.737%	9.035%	-0.298%
8	8.566%	8.757%	-0.191%
9	8.207%	8.500%	-0.293%
Total	100.000%	100.000%	0.000%
Maximum			1.396%
MAD			0.466%
MSE			0.004%

TABLE 1983 - ATTEMPT #1

NUMBER OF DIGITS IN THE 2ND POSITION, 142 FIELDS, 1983 PANEL DATA BASE, 19185 TAXPAYERS

Second Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
0	58103	52033	6070	11.665%	708.039
1	46578	49516	-2938	-5.933%	174.319
2	47349	47312	37	0.079%	0.029
3	43276	45360	-2084	-4.593%	95.706
4	43082	43612	-530	-1.215%	6.436
5	46059	42034	4025	9.577%	385.505
6	39415	40594	-1179	-2.906%	34.270
7	37984	39281	-1297	-3.303%	42.856
8	37243	38073	-830	-2.180%	18.086
9	35681	36955	-1274	-3.449%	43.951
Total	434770	434770	0	-2.257%	
Maximum			6070	11.665%	
MAD			2026	4.490%	
MSE			7168795	0.322%	
Total Chi-Squared					1509.196
Phi					0.05892
Contingency Coefficient					0.05882

TABLE 1983 - ATTEMPT #1

PERCENTAGE OF DIGITS IN THE 3RD POSITION FOR 142 VARIABLE FIELDS
ON THE 1983 PANEL DATA BASE FOR 19185 TAXPAYERS

Third Digit	Observed Percentage	Expected Percentage	Difference in Percentage
0	13.649%	10.178%	3.471%
1	9.274%	10.138%	-0.864%
2	10.113%	10.097%	0.016%
3	9.454%	10.057%	-0.603%
4	9.498%	10.018%	-0.520%
5	10.538%	9.979%	0.559%
6	9.229%	9.940%	-0.711%
7	9.700%	9.902%	-0.202%
8	9.474%	9.864%	-0.390%
9	9.071%	9.827%	-0.756%
Total	100.000%	100.000%	0.000%
Maximum			3.471%
MAD			0.809%
MSE			0.015%

TABLE 1983 - ATTEMPT #1

NUMBER OF DIGITS IN THE 3RD POSITION, 142 FIELDS, 1983 PANEL DATA BASE, 19185 TAXPAYERS

Third Digit	Observed Frequency	Expected Frequency	Difference in Frequency	Difference as a Percent of Expected	Chi-Squared of Frequency
0	55148	41123	14025	34.104%	4782.911
1	37469	40962	-3493	-8.527%	297.826
2	40862	40796	66	0.161%	0.106
3	38198	40635	-2437	-5.996%	146.096
4	38376	40477	-2101	-5.190%	109.047
5	42576	40319	2257	5.597%	126.303
6	37291	40162	-2871	-7.148%	205.204
7	39191	40008	-817	-2.043%	16.694
8	38280	39855	-1575	-3.951%	62.218
9	36651	39705	-3054	-7.692%	234.936
Total	404042	404042	0	-0.685%	
Maximum			14025	34.104%	
MAD			3269	8.041%	
MSE			24505340	1.460%	
Total Chi-Squared					5981.342
Phi					0.12167
Contingency Coefficient					0.12078